

Tau ID Status Report



Paul Padley
Algorithms Meeting April 28/00



Tasks and People

- Level 1
 - Andre, Yuri
- Level 2
 - Yuri, Bryan
- Level 3
 - Gustaaf
- PMCS
 - Bryan, Greg
- Reco
 - Qizhong
 - Leo, Dhiman, Marek (Multivariate stuff)
- Analysis
 - Qizhong, Leo, Naresh, Silke



Disclaimer

- Anything good in this talk should be credited to the people on the previous slide
- Any mistakes are my own fault, for preparing this talk in a hurry or as a result of my ignorance.
- These people are working hard and doing an excellent job
- More people participating would be more than welcome!

Recall Some Tau Basics

- ~85% decays are 1 prong
 - Ie. One charged particle plus any number of neutrals
- ~49% decays are 1 prong where the “prong” is a hadron
- 15% decays are 3 prong
- Tau reco is looking for the Hadronic decays
 - (lepton ID groups look for the leptons)
- So we look for narrow isolated jets as the basic tau objects

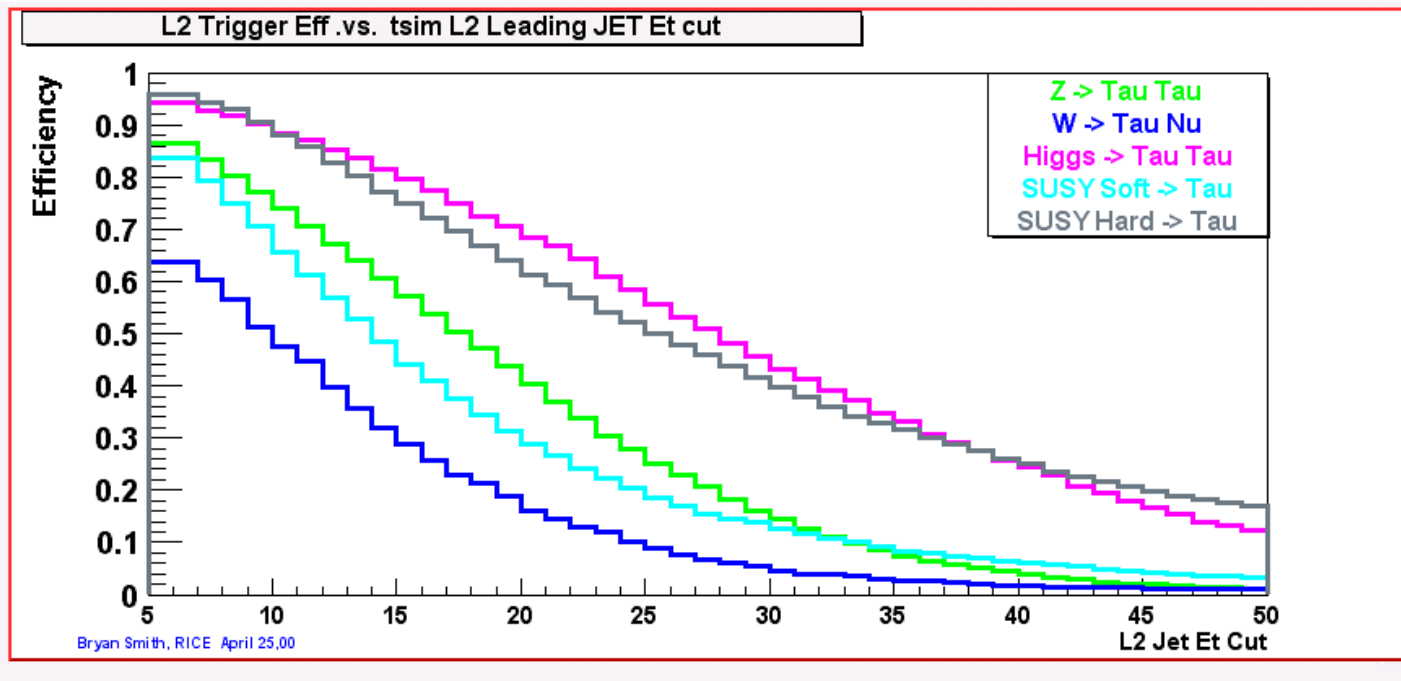
Level 1

- There are terms in the level 1 trigger list
- These are educated stabs at what should be done
- Some study has been made of tracking triggers, but Yuri has be pulled into hardware work
- A lot of work needs to be done

Level 2

- Again Yuri has done work, but has been pulled into hardware
- Bryan has begun to work on it

Example Plot

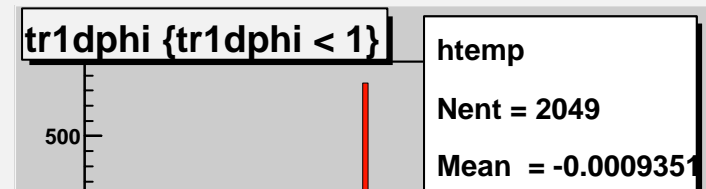
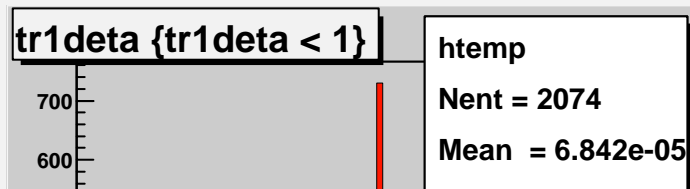
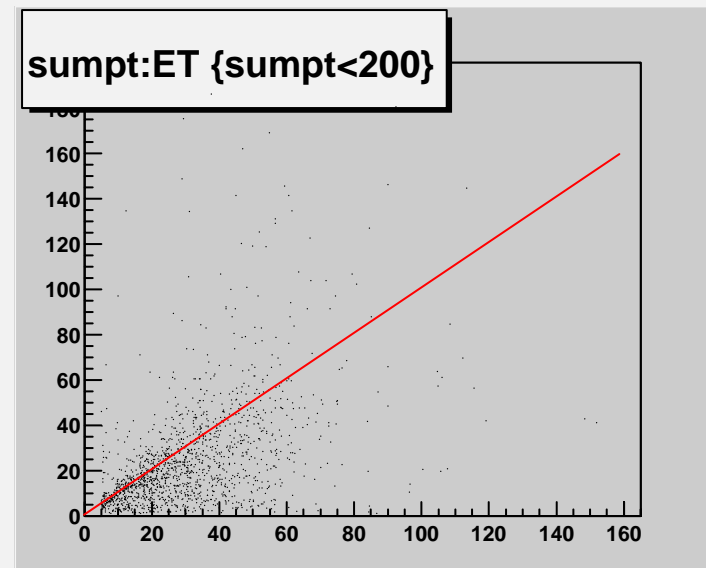
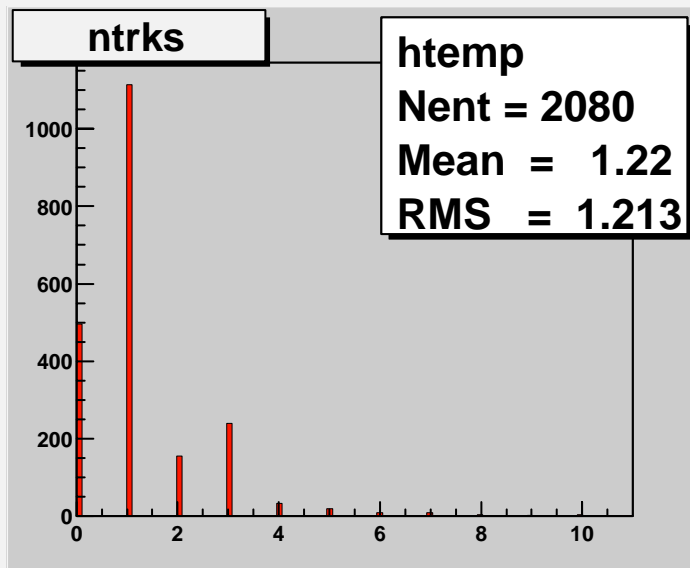


- This plot is only meant to demonstrate that some work is underway
- This work is in the earliest of stages and is hampered by a lack of tools



Level 3

- Level 3 effort is moving along well – mainly by Gustaaf
- Recently has been trying the level 3 stiff tracker and it seems to work very well



TauReco Selection

- starts from a calorimeter cluster (CalCone07, $E_T > 8 \text{ GeV}$)
- $rms < 0.25$
- $em \text{ fraction} < 0.95$
- $1 \leq \text{no. of matched tracks} \leq 5$ (within 0.2×0.2 window of cluster in η - ϕ space. $p_T \geq 0.5 \text{ GeV}/c$)
- NOTE: the point of this selection is not to get pure taus, but a reasonable set of candidates

Note Variables

♦ **RMS** (width of the jet):

$$RMS = \sqrt{\phi_{width}^2 + \eta_{width}^2} < 0.25$$

$$\phi_{width} = \sqrt{\frac{\sum (\phi_i - \phi)^2 E_T^i}{E_T}}$$

$$\eta_{width} = \sqrt{\frac{\sum (\eta_i - \eta)^2 E_T^i}{E_T}}$$

i is the cells above threshold (100 MeV)



Additional Useful Variable

♦ *Profile:*

$$P = \frac{E_{T1} + E_{T2}}{E_T}$$

E_{T1} and E_{T2} are the two highest E_T towers

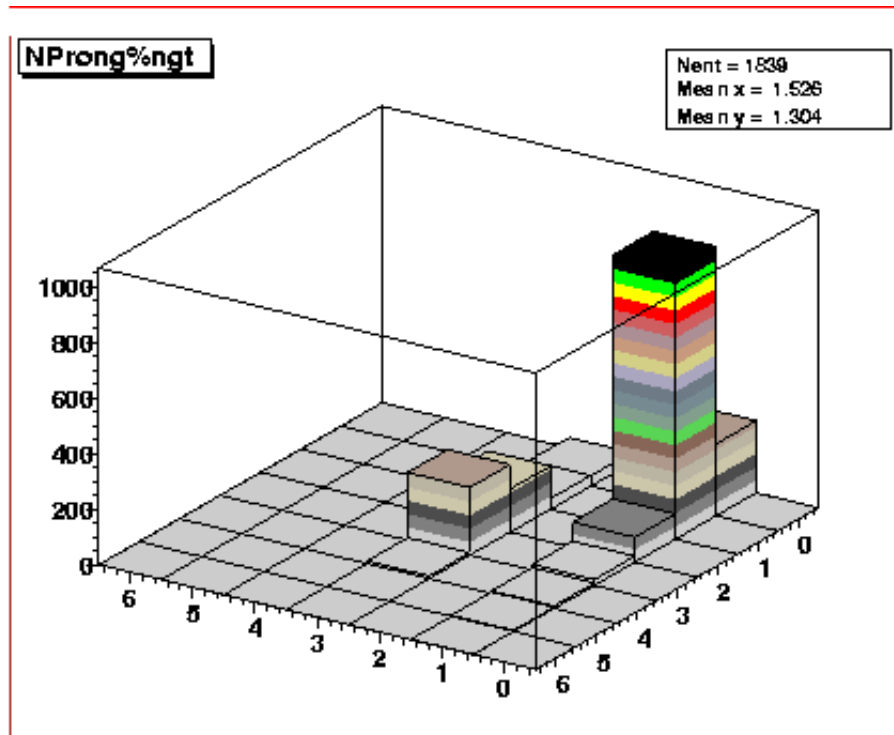
E_T is the total E_T

Tau jets have a larger value of P than QCD jets



Taus and Tracking

- Tracking in taus is being studied by Silke



Efficiency

denominator:

number of generated taus that have:

- decayed hadronically
- *visible* ($\nu_\tau E_T$ subtracted) $E_T > 8 \text{ GeV}$
- at least one charged pion with $p_T > 0.5 \text{ GeV}/c$
- at least one associated ($dR < 0.2$) calorimeter cluster
(here dR uses the ν_τ subtracted direction of the tau)

numerator:

number of generated taus in denominator
that have an associated tau candidate
after a specific set of cuts.

Definitions of some of the terms to be used

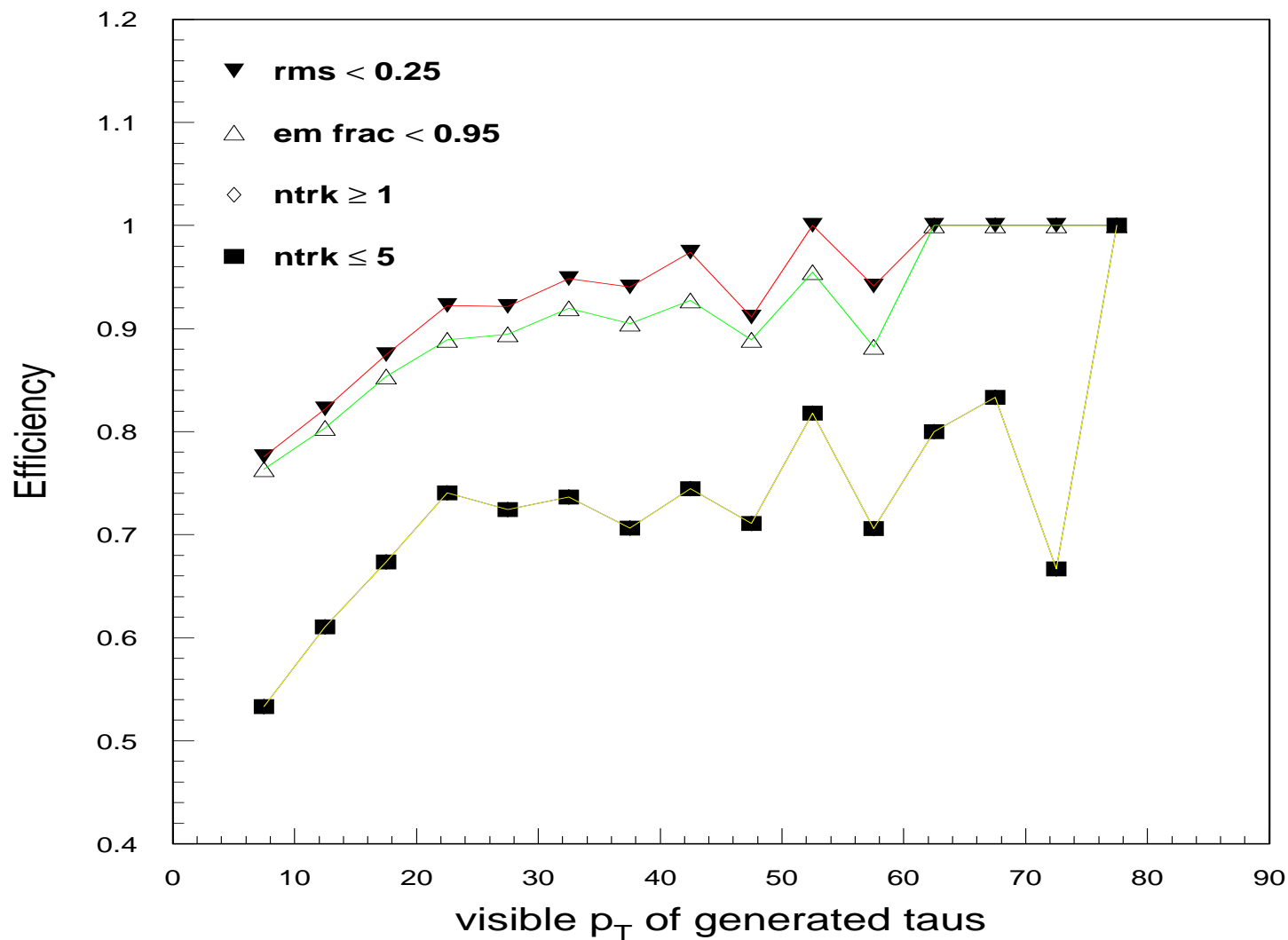
- $dR = \sqrt{\{(\Delta\eta)^2 + (\Delta\phi)^2\}}$
i.e. distance in η - ϕ space
- association: $dR < 0.2$
- visible ‘stuff’: tau neutrino ‘stuff’
subtracted. ‘Stuff’ can be p_T , η , etc..

Monte Carlo Samples for Efficiency Study

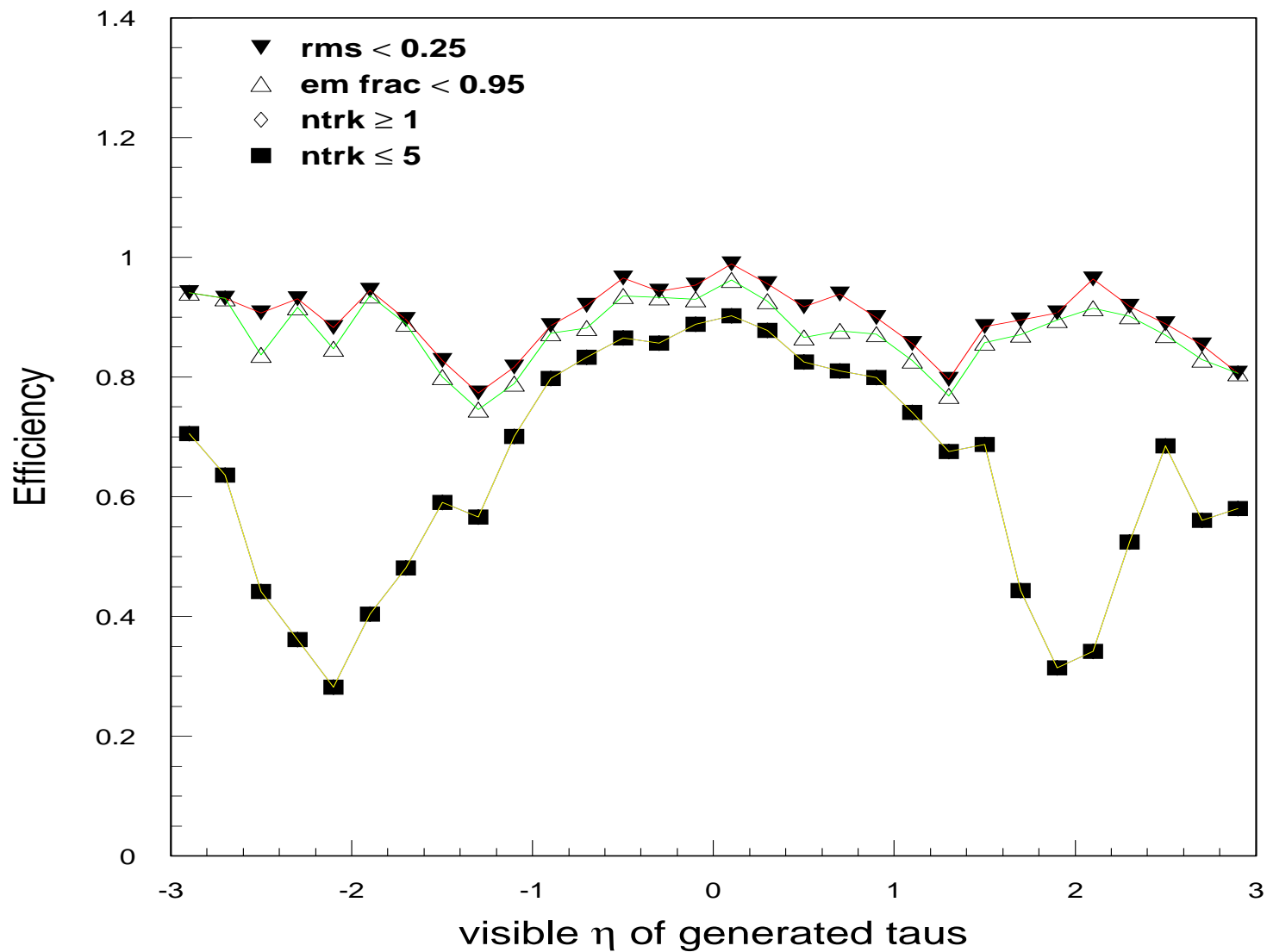
- pythia $Z \rightarrow \tau\tau$, QCD
- pmc02.01, preco03.07.00
- average 1.1 minbias



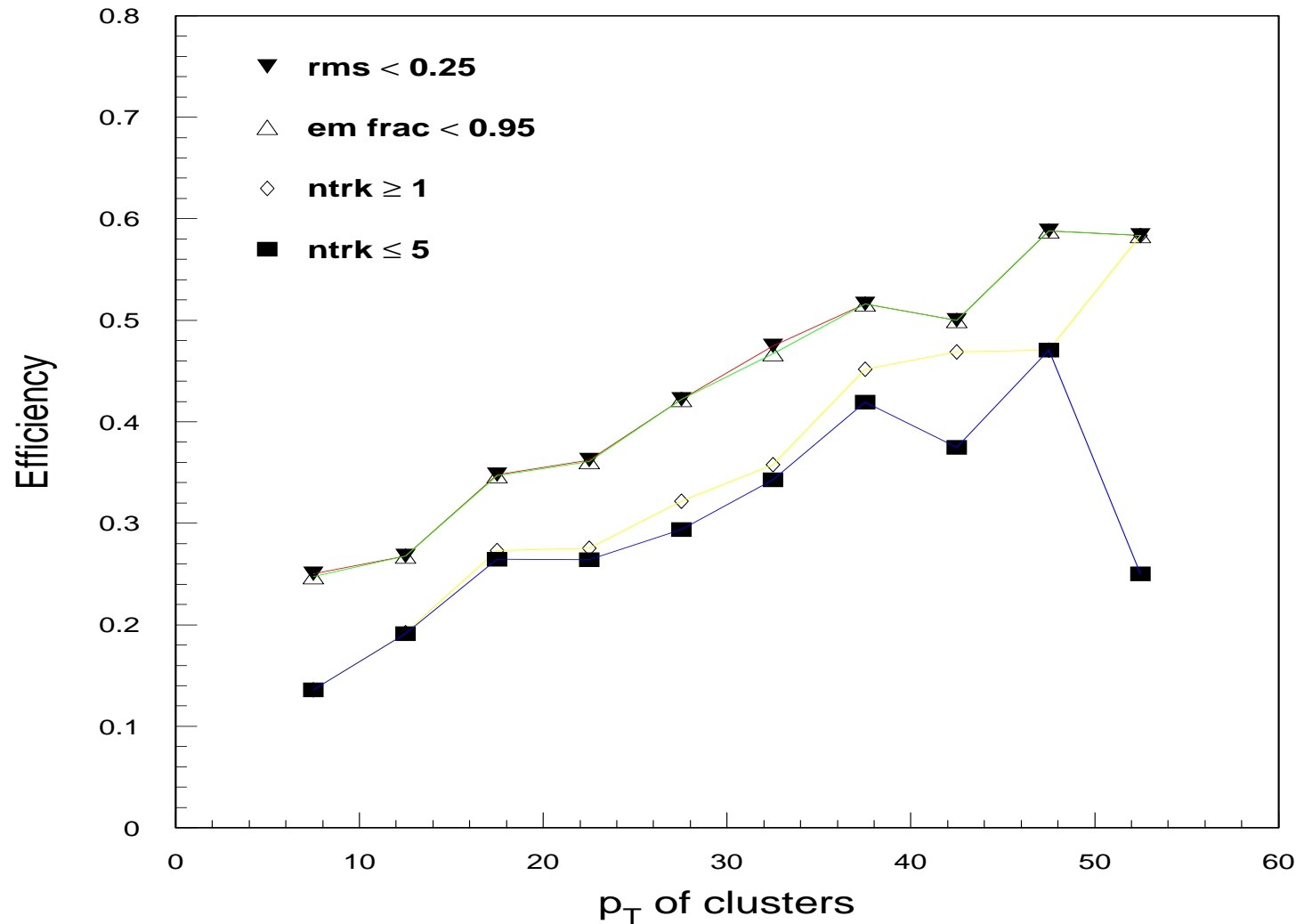
$Z \rightarrow \tau\tau$, pmc02.01, preco03.07.00, pythia, mb1.1av



$Z \rightarrow \tau\tau$, pmc02.01, preco03.07.00, pythia, mb1.1av



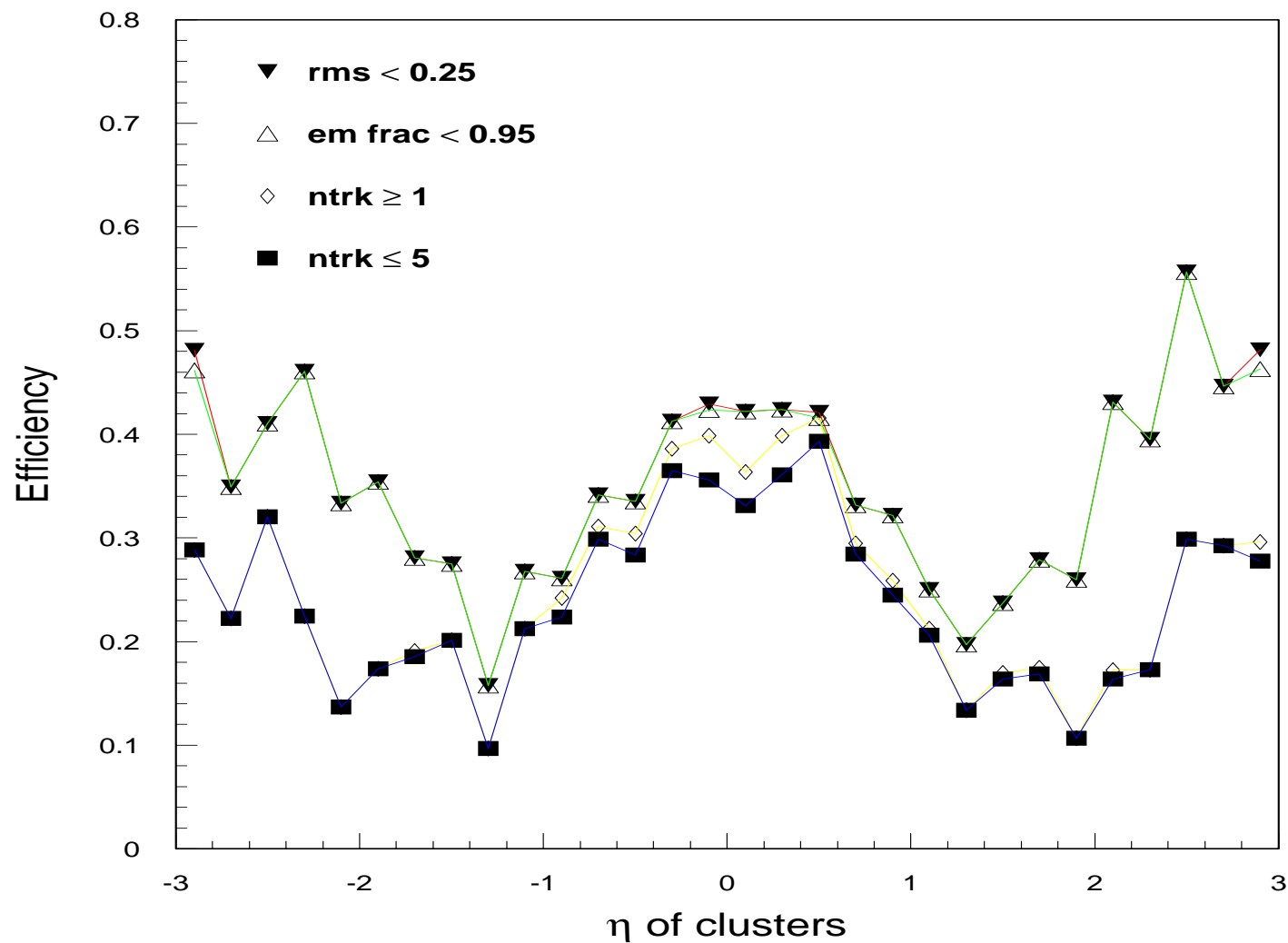
QCD $p_T > 20$, pmc02.01, preco03.07.00, pythia, mb1.1av



* Here the denominator is simply the total no. of calorimeter clusters.



QCD $p_T > 20$, pmc02.01, preco03.07.00, pythia, mb1.1av



Efficiencies

	$Z \rightarrow \tau\tau$	
generated	3760(7520 τ)	
hadronic	4818	
kinematic cuts	4177	
associated cluster	3812	
Selection Cuts	Single Pass	Cumulative
$rms < 0.25$	90.3%	90.3%
$emf < 0.95$	97.3%	87.6%
$N(track) \geq 1$	78.9%	70.1%
$N(track) < 5$	99.97%	70.1%

★ The 2nd - 4th rows show the the no. of generated taus passing (cumulatively) each cut. The lower section shows the efficiencies. Single pass means that only that particular cut is applied.



Multivariate Tools

- In order to select a more pure sample of taus Multivariate techniques are used
- Recall that in Run I an H-Matrix was used
- That is being done again and is implemented in Taureco

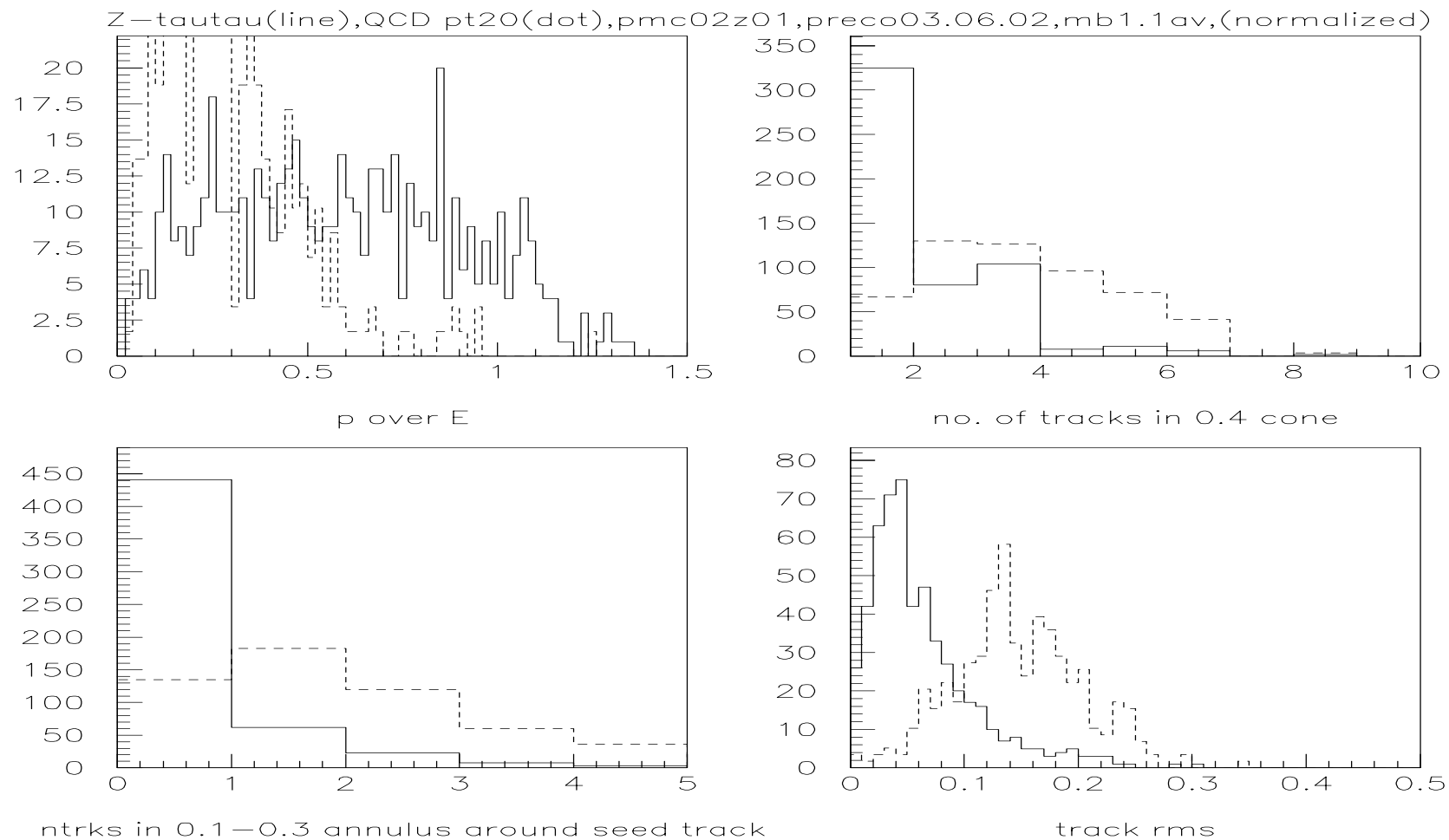
Work on improving the Hmatrix is continuing, which is shown next

Training Sample

- Signal: 600 $Z \rightarrow \tau\tau$ events with no associated jets
- Background: 600 pythia QCD pt20
- pmc02.01, preco03.06.02
- Minbias: 1.1average
- $-1 \leq \eta$ of cluster ≤ 1

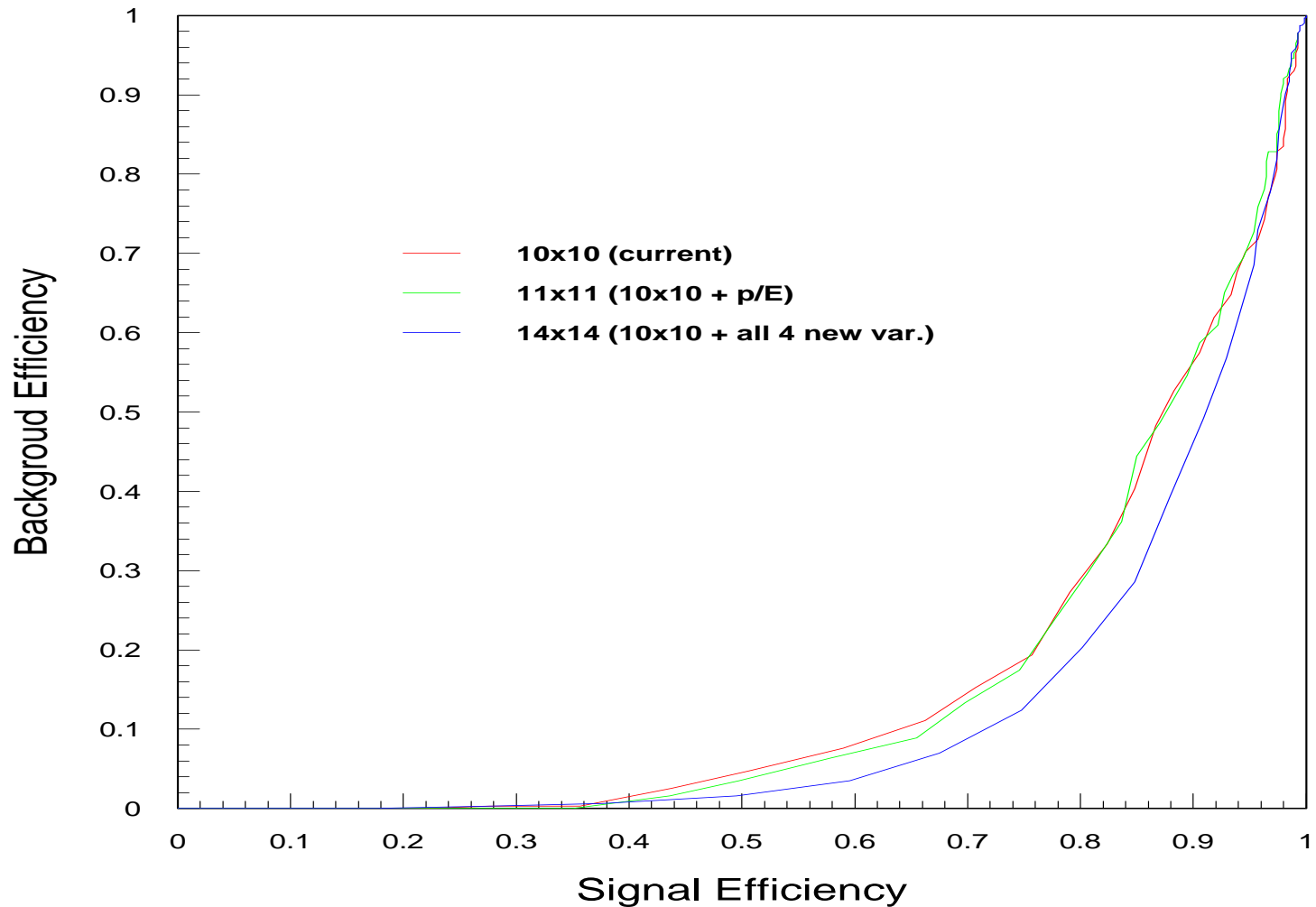
New H-matrix Variables

- p/E : momentum of seed track (highest p_T matched track)/tot. energy of cluster
- track rms (within 0.4 cone of cluster)
- no. of tracks in 0.4 cone around cluster
- no. of tracks in 0.1-0.3 annulus around seed track (isolation)



Efficiencies by cutting on Fisher variable

pmc02.01, preco03.06.02, pythia, mb1.1av



Neural Network

- Recently a Neural Network has been tried
- The initial result was promising

NN variables

- E_{EM1}/E_{total}
- E_{EM4}/E_{total}
- $E_{3 \times 3}/E_{total}$
- $E_{5 \times 5}/E_{total}$
- E_{hot2}/E_{total}
- $\log_{10}(E_{total})$
- This is a subset of the “standard” Hmatrix variables

Samples

- Samples used for training:
 - Signal: $Z \rightarrow \tau \tau + 0 \text{ jet} + 1.1 \text{ (avg) min-bias}$, 2000 events
 - Background: $\text{QCD (} p_T > 20 \text{ GeV)} + 1.1 \text{ (avg) minbias}$, 1000 events
- Samples used for testing:
 - $Z \rightarrow \tau \tau + 0 \text{ jet} + 1.1 \text{ (avg) minbias}$, 600 events
 - $\text{QCD (} p_T > 20 \text{ GeV)} + 1.1 \text{ (avg) min-bias}$, 1000 events
 - $Z \rightarrow \tau \tau + 0 \text{ jet} + 0 \text{ minbias}$, 1000 events
 - Single tau, $p_T = 50 \text{ GeV}$, 0 minbias, 2000 events

The Neural Network

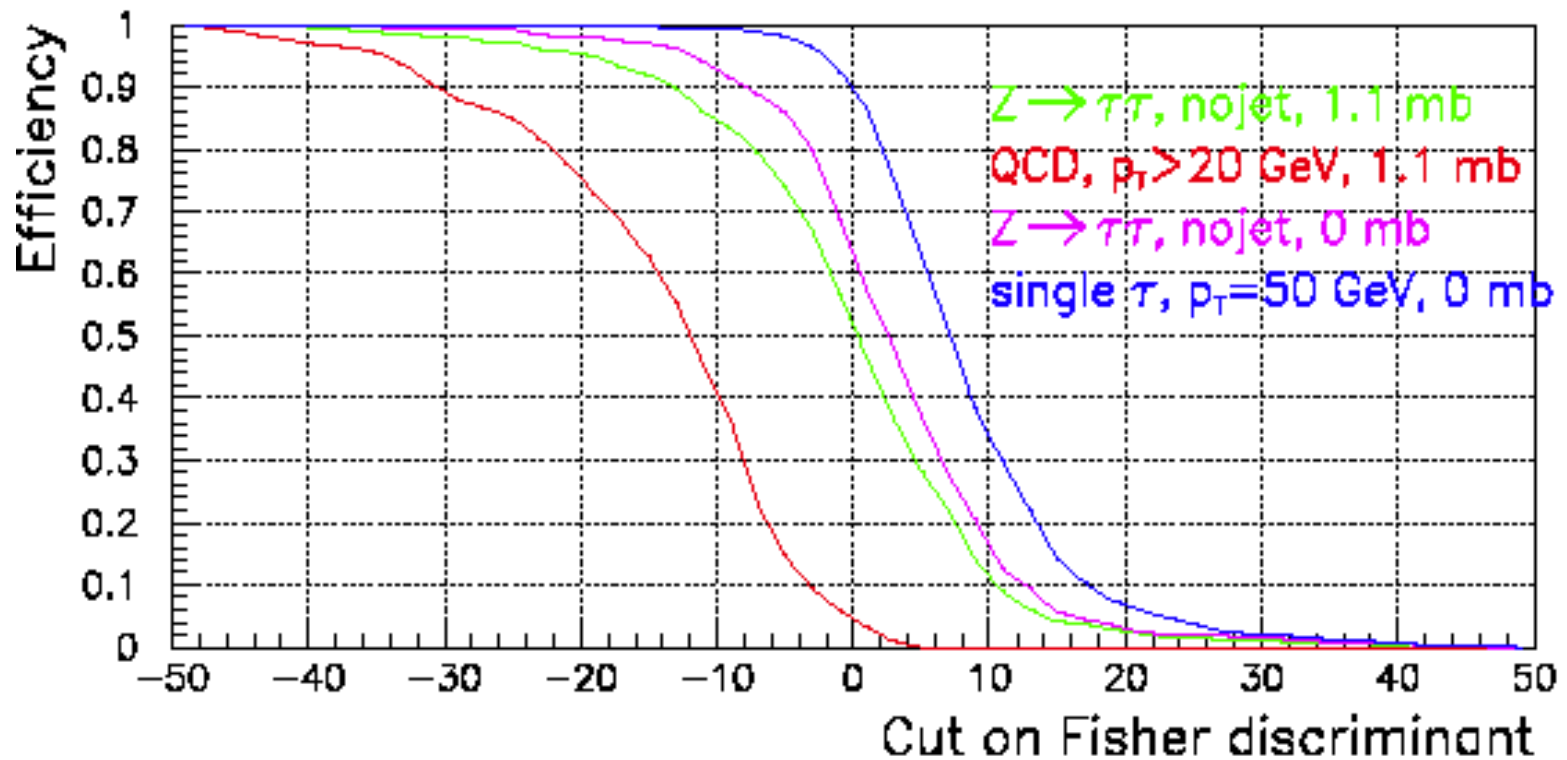
- *MLPfit* (MultiLayer Perceptrons)

<http://home.cern.ch/schwind/code.html>

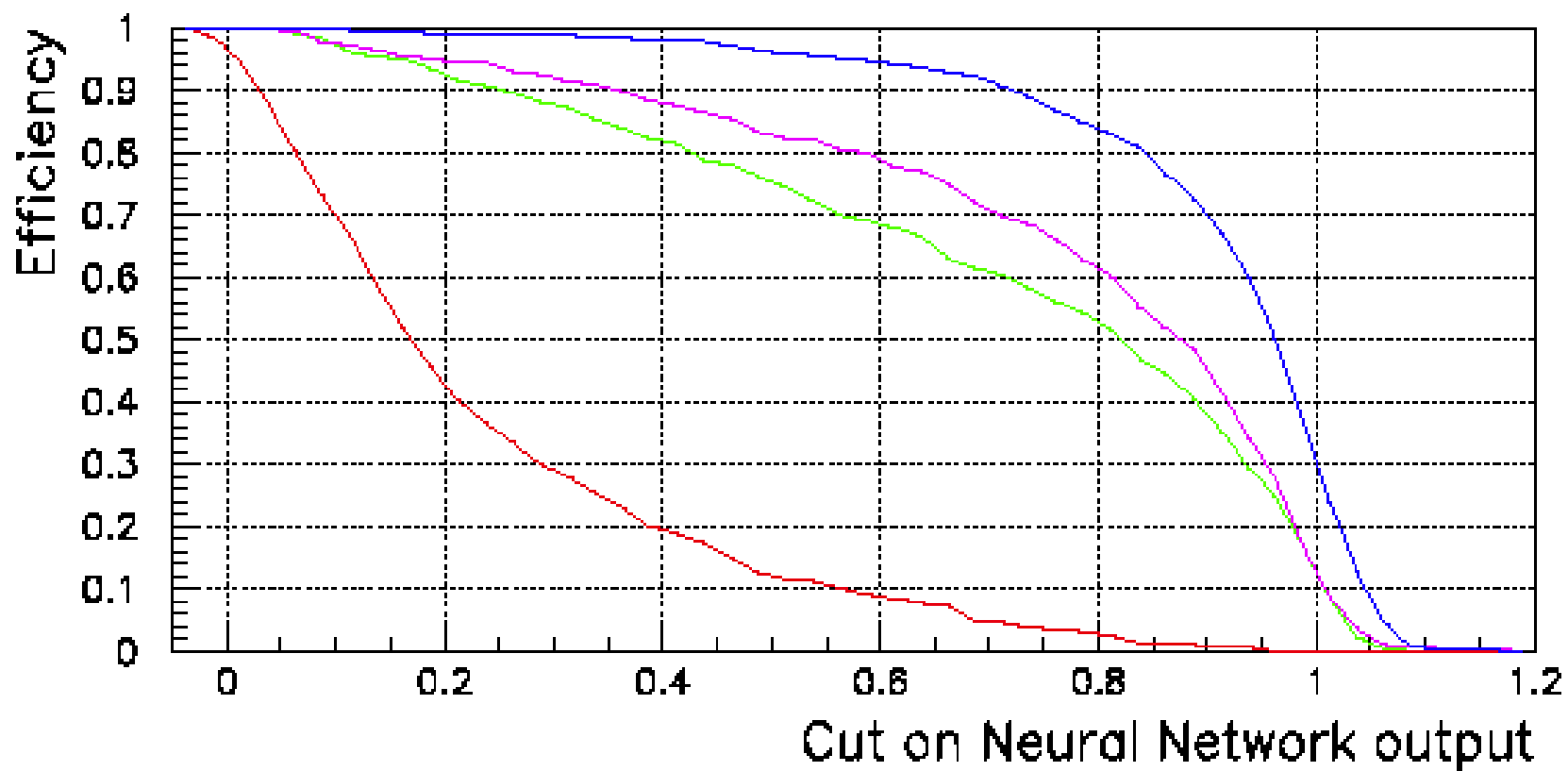
- Used in PAW with default parameter set
- 6 input nodes, 1 hidden layer with 12 nodes, 1 output node
- Method of minimization: line search (BFGS)
- Number of training iterations (epochs): 200



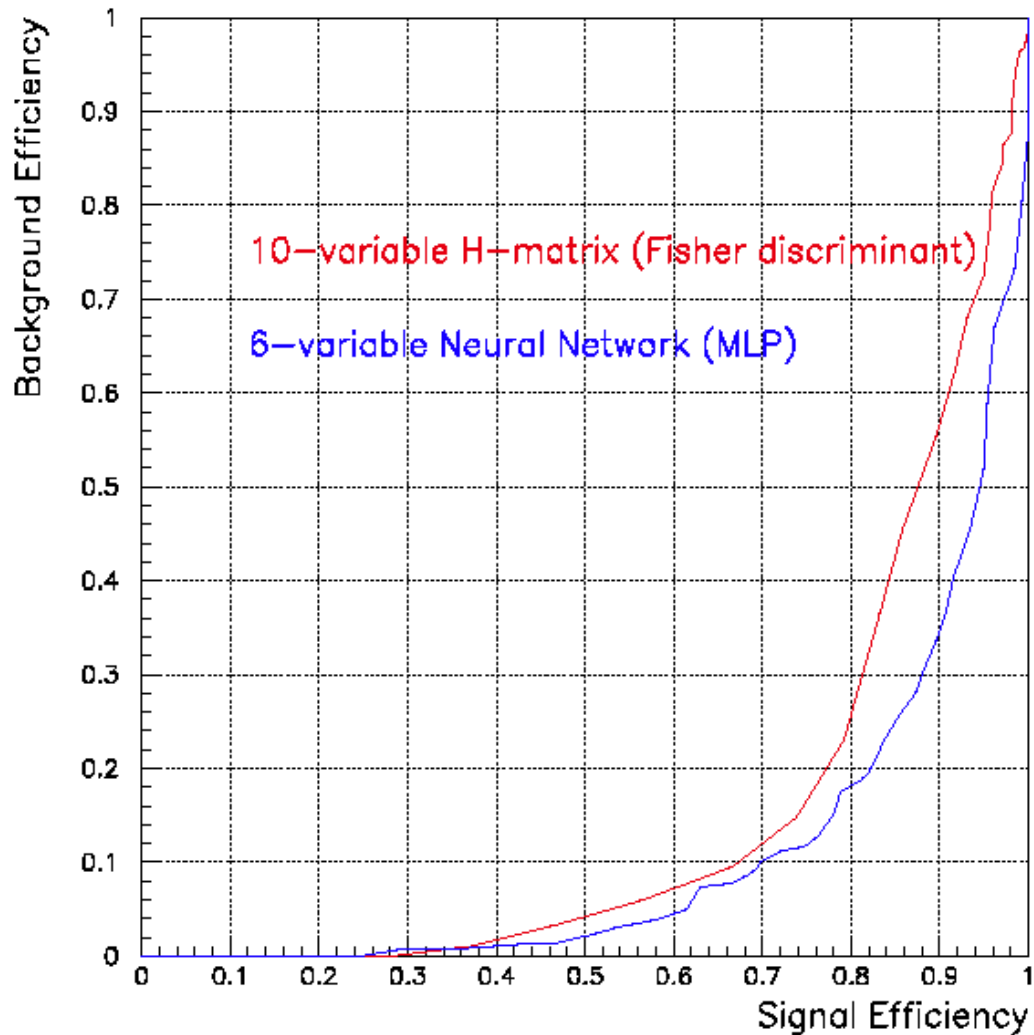
HMatrix cut



NN cut



Comparison



Clear Improvement

- Previous NN slides were made a few weeks ago.
- There was no optimization, just a first attempt at it, but it showed an improvement
- This week Dhiman showed new results with a different set of variables

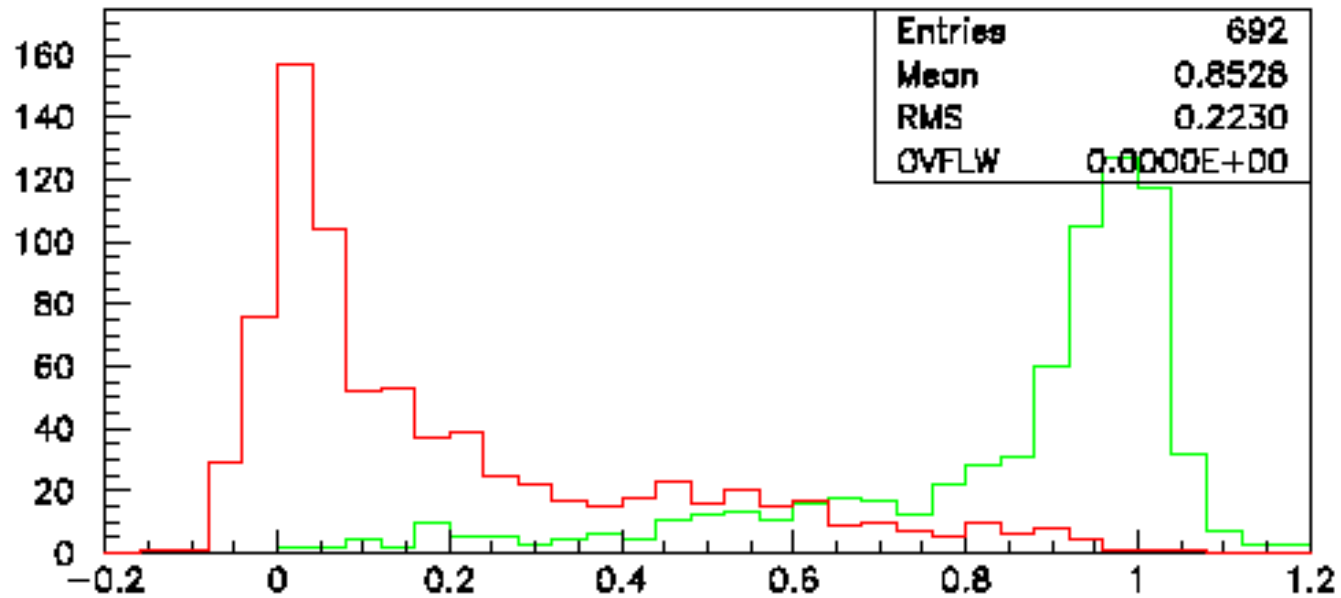


Another set of NN variables

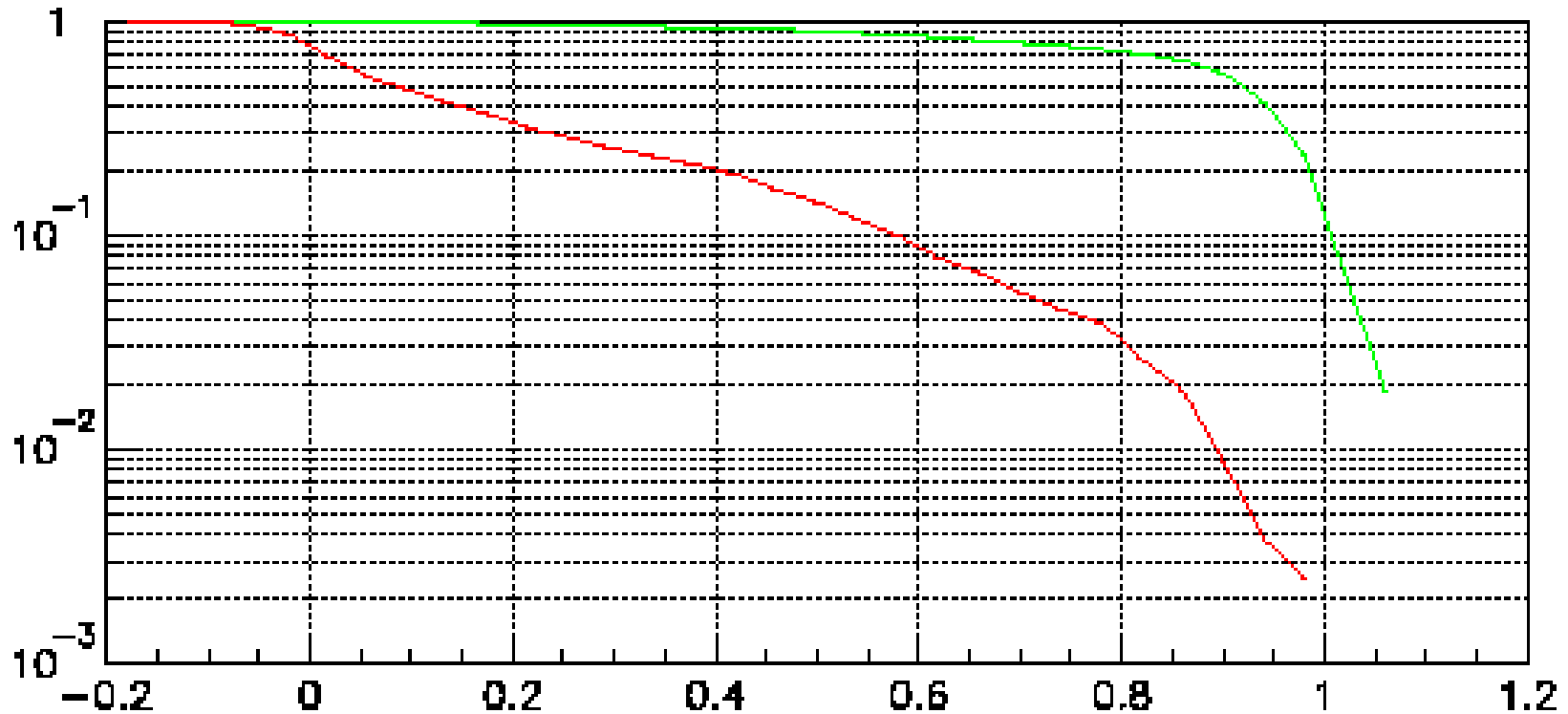
- E_{EM3}/E_{total}
- E_{EM4}/E_{total}
- E_{FH}/E_{total}
- RMS
- E_{hot2}/E_{total}
- $\log_{10}(E_{total}) - 1$
- $\log_{10}(N_{0.5} + 1)$
- $(N_{0.5} - N_{0.2}) / (N_{0.5} + N_{0.2} + 1)$
- **invariant mass** $\sqrt{(cal\ cluster\ E)^2 - P^2(\text{ tracks in } 0.5\ cone)}$



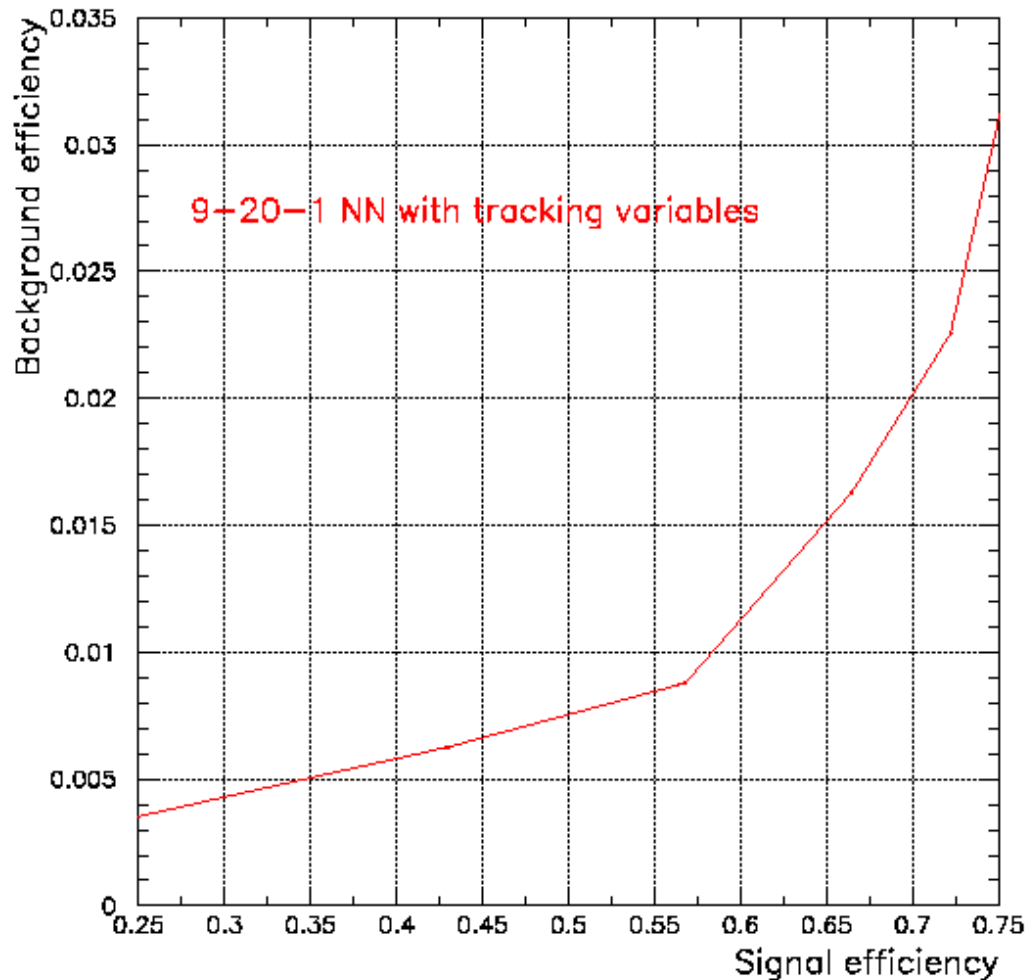
NN Output



NN Efficiency



Factor of 5 Improvement



Conclusions

- Tau reco is stable
 - But still there is work to be done,
 - Need person to take responsibility for tau's in the forward direction
- Multivariate analysis is promising and lots of work going on
- The Tau ID group does not suffer from a surplus of people.

